

Shrinking the Quadratic Estimator

Ethan Anderes* and Debashis Paul†

Statistics Department

University of California, Davis, CA 95616

We study a regression characterization for the quadratic estimator of weak lensing, developed by Hu and Okamoto [16, 17, 29], for cosmic microwave background observations. This characterization motivates a modification of the quadratic estimator by an adaptive Wiener filter which uses the robust Bayesian techniques described in [3, 4, 36]. This technique requires the user to propose a fiducial model for the spectral density of the unknown lensing potential but the resulting estimator is developed to be robust to misspecification of this model. The role of the fiducial spectral density is to give the estimator superior statistical performance in a “neighborhood of the fiducial model” while controlling the statistical errors when the fiducial spectral density is drastically wrong. Our estimate also highlights some advantages provided by a Bayesian analysis of the quadratic estimator.

I. INTRODUCTION

The cosmic microwave background (CMB) measures temperature fluctuations in the early Universe some 400,000 years after the big bang. These fluctuations provide us with a picture of the Universe at the instant of recombination and contains a wealth of information for cosmology and cosmic structure. One important characteristic of the observed CMB is that the photon paths have been distorted, or lensed, from the gravitational effect of intervening matter. Estimating this lensing is important for a number of reasons including, but not limited to, understanding cosmic structure, constraining cosmological parameters [20, 34] and detecting gravity waves [21, 23, 30]. There have been a number of proposed estimators for the lensing of the CMB [1, 5, 6, 12, 13, 16, 17, 29]. The most widely used estimate was developed in [16, 17, 29] and is referred to as ‘the quadratic estimator’. This estimator, up to leading order, is an unbiased minimum variance estimator. In this paper we study the potential advantages obtained by relaxing the unbiasedness constraint, borrowing statistical techniques from Bayesian and regression theory.

The effect of lensing is to simply remap the CMB, preserving surface brightness. Up to leading order, the remapping displacements are given by $\nabla\phi$, where ϕ denotes a lensing potential and is the planer projection of a three dimensional gravitational potential (see [7], for example). The lensed CMB can then be written $\Theta(\mathbf{x} + \nabla\phi(\mathbf{x}))$ where $\Theta(\mathbf{x})$ denotes the unlensed CMB temperature fluctuations projected to the observable sky. In this paper we work in the small angle limit and use a flat sky approximation so that $\mathbf{x} \in \mathbb{R}^2$. The isotropic and Gaussian assumptions for $\Theta(\mathbf{x})$ translates to independence under the Fourier transform. However, for a fixed lensing potential ϕ , the lensed CMB becomes non-isotropic, leading to correlated Fourier modes. The quadratic estimator takes advantage of this correlation

and uses weighted sums of Fourier cross products to unbiasedly (up to leading order) estimate the lensing potential.

By relaxing the unbiasedness constraint and utilizing robust Bayesian techniques (originally developed in [3, 36] for statistical regression) we propose an adaptive shrinkage adjustment to the nominal quadratic estimator. The new estimator is given by

$$\tilde{\phi}(\mathbf{L}) = \left[1 - F_L^{\phi} \frac{2A_L}{C_{L,\text{fid}}^{\phi\phi} + 2A_L} \right] \hat{\phi}(\mathbf{L}) \quad (1)$$

where $\hat{\phi}(\mathbf{L})$ is the regular quadratic estimator at frequency $\mathbf{L} \in \mathbb{R}^2$; $C_{L,\text{fid}}^{\phi\phi}$ is a fiducial spectral density model for the gravitational potential ϕ ; $2A_L$ is the approximate variance of $\hat{\phi}(\mathbf{L})$ derived in [17]; and F_L^{ϕ} is an adaptive shrinkage factor defined in Section IV B. The formula makes it clear that the estimator is essentially an adaptive Wiener filter: shrinking $\hat{\phi}(\mathbf{L})$ when $2A_L/C_{L,\text{fid}}^{\phi\phi}$ is large and retaining $\hat{\phi}(\mathbf{L})$ when $2A_L/C_{L,\text{fid}}^{\phi\phi}$ is small.

The adaptive shrinkage factor F_L^{ϕ} is derived as a posterior expected value in a robust Bayesian procedure and adapts the Wiener filter to account for uncertainty associated with misspecification of $C_{L,\text{fid}}^{\phi\phi}$. Indeed, the Bayesian viewpoint is the principal advantage of the estimate: using posterior draws of the lensing potential ϕ , one can construct estimates and quantify the uncertainty for any non-linear function of the gravitational potential, including spectral density estimate. Moreover, in Section IV B we show that the robust Bayesian procedure is easy to simulate without resorting to expensive Markov Chain Monte Carlo (MCMC).

The non-informative prior for this Bayesian procedure is parameterized by a fiducial spectral density $C_{L,\text{fid}}^{\phi\phi}$ for the lensing potential ϕ and utilizes a hierarchical structure to induce robustness. Robustness, in this context, pertains to the stability under mis-specification of the fiducial spectral density. We illustrate this property with simulations that demonstrate the similarity between $\tilde{\phi}(\mathbf{L})$ and the optimal Wiener filter if one had access to the true spectral density for ϕ , even when the fiducial

* Supported by NSF grant 1007480.

† Supported by NSF grants DMR-1035468, DMS-1106690.

model is wrong by a relatively large factor (similar robustness properties have been demonstrated in [3]). In fact, $\tilde{\phi}(\mathbf{L})$ also has a frequentist interpretation that is independent of the Bayesian underpinnings. At the end of Section IV B we discuss the connection with a generalized James-Stein shrinkage estimator [18], where the role of the fiducial spectral density is essentially to specify a shrinkage direction.

It should be noted that our Bayesian posterior is essentially an approximation since we model the lensing operation using the same first order Taylor approximation present in the quadratic estimator. Therefore, the well known bias in the quadratic estimator, derived in [10, 22], is still present in the estimate $\tilde{\phi}(\mathbf{L})$ and in the associated posterior samples. In the context of spectral density estimation one may be able to simply subtract this bias using the fiducial model $C_{L,\text{fid}}^{\phi\phi}$. However this would require some stability of the bias under misspecification of $C_{L,\text{fid}}^{\phi\phi}$ which is yet to be established.

The remainder of this paper is organized as follows. In Section II we establish the connection between regression and the quadratic estimator. This connection is then used in Section III to derive the optimal Wiener filter of the quadratic estimate when the true spectral density for ϕ is known. In Section IV we derive the adaptive shrinkage estimator (1) using Bayesian techniques when the spectral density for ϕ is unknown. Finally, in Section V we present some simulation examples exploring the advantages of the estimator (1) and its Bayesian interpretation.

II. REGRESSION AND THE QUADRATIC ESTIMATOR

In this section we characterize the quadratic estimator in terms of generalized least squares regression. The main utility of this connection is the incorporation of tools from statistical regression that lead to the shrinkage estimators given in sections III and IV.

The quadratic estimator is derived under the assumption that the observed lensed CMB field is contaminated by additive noise and an instrumental beam. Throughout this paper we let $\tilde{\Theta}(\mathbf{x})$ denote the observed CMB field with lensing, beam (denoted by φ) and noise (denoted $N(\mathbf{x})$) so that

$$\tilde{\Theta}(\mathbf{x}) = N(\mathbf{x}) + \int \Theta(\mathbf{y} + \nabla\phi(\mathbf{y}))\varphi(\mathbf{x} - \mathbf{y}) d\mathbf{y}.$$

The quadratic estimator is based on a first order Taylor approximation in $\nabla\phi$ on the lensed CMB field: $\Theta(\mathbf{x} + \nabla\phi(\mathbf{x})) = \Theta(\mathbf{x}) + \nabla\phi(\mathbf{x}) \cdot \nabla\Theta(\mathbf{x}) + O(\phi^2)$. By truncating $O(\phi^2)$ terms one gets the following approximation for the correlation of Fourier modes induced by lensing:

$$\langle \tilde{\Theta}(\ell + \mathbf{L})\tilde{\Theta}(\ell)^* \rangle \approx \delta_{\mathbf{L}} C_{\ell,\text{expt}}^{\Theta\Theta} + f_{\mathbf{L}}(\ell)\phi(\mathbf{L}) \quad (2)$$

where $f_{\mathbf{L}}(\ell) \equiv (2\pi)^{-1}[(\ell + \mathbf{L}) \cdot \mathbf{L} C_{\ell+\mathbf{L}}^{\Theta\Theta} - \ell \cdot \mathbf{L} C_{\ell}^{\Theta\Theta}] \varphi(\ell + \mathbf{L})\varphi^*(\ell)$, $C_{\ell}^{\Theta\Theta}$ denotes the flat sky power spectrum for Θ and $C_{\ell,\text{expt}}^{\Theta\Theta} \equiv |\varphi(\ell)|^2 C_{\ell}^{\Theta\Theta} + C_{\ell}^{NN}$ so that $C_{\ell,\text{expt}}^{\Theta\Theta}$ denotes the power spectrum for the unlensed CMB corrupted with experimental noise and beam. To avoid any confusion notice that we are adopting the notational convention of [27] (see Section 4.1) so that $C_{\ell}^{\Theta\Theta}$ is defined by $\langle \Theta(\ell)\Theta(\ell')^* \rangle = \delta_{\ell-\ell'} C_{\ell}^{\Theta\Theta}$ where $\Theta(\ell) \equiv \int e^{-i\mathbf{x}\cdot\ell} \Theta(\mathbf{x}) \frac{d^2\mathbf{x}}{2\pi}$ and $\delta_{\ell} \equiv \int e^{i\mathbf{x}\cdot\ell} \frac{d^2\mathbf{x}}{(2\pi)^2}$. Throughout the remainder of the paper we let angled brackets (as used in (2)) denote expectation over the unlensed CMB and the instrumental noise, while holding the potential ϕ fixed.

Equation (2) exposes the nonstationarity in the lensed temperature field through the cross correlation of Fourier modes. The quadratic estimator uses this correlation to estimate $\phi(\mathbf{L})$ by weighted averaging cross products of $\tilde{\Theta}(\ell)$ separated at lag $\mathbf{L} \neq 0$. In particular, let $k = 1, 2, \dots$ index a set of frequencies $\ell_k \in \mathbb{R}^2$. Now, for each k define the normalized cross product at lag \mathbf{L} , $\mathbf{v}_{L,k} \equiv \tilde{\Theta}(\ell_k + \mathbf{L})\tilde{\Theta}(\ell_k)^*/f_{\mathbf{L}}(\ell_k)$ and the corresponding vector of these cross products $\mathbf{v}_{\mathbf{L}} \equiv (\mathbf{v}_{L,1}, \mathbf{v}_{L,2}, \dots)^{\dagger}$. Since $\mathbf{L} \neq 0$, equation (2) implies that each $\mathbf{v}_{L,k}$ is a noisy unbiased estimate of $\phi(\mathbf{L})$ (up to the approximation in (2)). Writing this statement in a regression setting one obtains

$$\mathbf{v}_{\mathbf{L}} = \mathbf{1}\phi(\mathbf{L}) + \mathbf{e}_{\mathbf{L}} \quad (3)$$

where for each fixed \mathbf{L} , $\mathbf{e}_{\mathbf{L}}$ is an error vector and $\mathbf{1}$ is a vector of ones. Approximation (2) establishes that $\langle \mathbf{e}_{\mathbf{L}} \rangle \approx 0$. The generalized least squares regression estimator for $\phi(\mathbf{L})$ is then

$$\hat{\phi}(\mathbf{L}) \equiv (\mathbf{1}^{\dagger} \mathbf{N}_{\mathbf{L}}^{-1} \mathbf{1})^{-1} \mathbf{1}^{\dagger} \mathbf{N}_{\mathbf{L}}^{-1} \mathbf{v}_{\mathbf{L}} \quad (4)$$

where $\mathbf{N}_{\mathbf{L}}$ is the covariance matrix of $\mathbf{e}_{\mathbf{L}}$ which is approximated as follows:

$$\begin{aligned} (\mathbf{N}_{\mathbf{L}})_{k,j} &\equiv \langle \mathbf{v}_{L,k} \mathbf{v}_{L,j}^* \rangle - |\phi(\mathbf{L})|^2 \\ &\approx [\delta_{\ell_k - \ell_j}^2 + \delta_{\ell_k + \ell_j + \mathbf{L}}^2] \frac{C_{\ell_k + \mathbf{L}, \text{expt}}^{\Theta\Theta} C_{\ell_j, \text{expt}}^{\Theta\Theta}}{f_{\mathbf{L}}(\ell_k) f_{\mathbf{L}}(\ell_j)^*}. \end{aligned}$$

The above approximation is obtained from (2) using Wick's theorem and discarding any $O(\phi)$ terms. Also notice that, in practice, the covariance matrix $\mathbf{N}_{\mathbf{L}}$ is based on a discrete grid approximation in Fourier space that arises from finite sky observations of $\tilde{\Theta}$. In particular, δ_{ℓ} is approximated as $1/\Delta\mathbf{L}$ when $\ell = 0$ and zero otherwise, where $\Delta\mathbf{L}$ is the area element of the grid in Fourier space. For the remainder of the paper we do not distinguish the discrete case versus the continuous and simply equate δ_0 with $1/\Delta\mathbf{L}$ leaving it understood that equality holds in the limit as $\Delta\mathbf{L} \rightarrow 0$.

This is not the typical derivation of the quadratic estimator. However, it should be no surprise they are related since both are minimum variance estimators. To connect the two, notice that the only reason $\delta_{\ell_k + \ell_j + \mathbf{L}}$ appears in $\mathbf{N}_{\mathbf{L}}$ is that the terms $\tilde{\Theta}(\ell + \mathbf{L})\tilde{\Theta}(\ell)^*/f_{\ell}(\mathbf{L})$ and

$\tilde{\Theta}(\ell' + \mathbf{L})\tilde{\Theta}(\ell')^*/f_{\ell'}(\mathbf{L})$ are the same when $\ell + \ell' + \mathbf{L} = 0$. Therefore, if we only include unique observations in $\mathbf{v}_{\mathbf{L}}$, $\mathbf{N}_{\mathbf{L}}$ becomes diagonal. It is now easy to see that

$$\mathbf{1}^\dagger \mathbf{N}_{\mathbf{L}}^{-1} \mathbf{v}_{\mathbf{L}} = \frac{1}{2\delta_0} \int \frac{f_{\mathbf{L}}(\ell)^* \tilde{\Theta}(\ell + \mathbf{L}) \tilde{\Theta}(\ell)^*}{C_{\ell+\mathbf{L},\text{expt}}^{\Theta\Theta} C_{\ell,\text{expt}}^{\Theta\Theta}} d\ell \quad (5)$$

$$\mathbf{1}^\dagger \mathbf{N}_{\mathbf{L}}^{-1} \mathbf{1} = \frac{1}{2\delta_0} \int \frac{|f_{\mathbf{L}}(\ell)|^2}{C_{\ell+\mathbf{L},\text{expt}}^{\Theta\Theta} C_{\ell,\text{expt}}^{\Theta\Theta}} d\ell. \quad (6)$$

where the equalities are understood to hold in the continuous-to-discrete approximation inherent in δ_0 and the above integrals. Notice that the factor of $1/2$ comes from the fact that the integrals have non-unique terms in the integrand. Returning to the original characterization of the quadratic estimator developed in [16, 17], $\hat{\phi}$ has the form $A_{\mathbf{L}} \int_{\mathbb{R}^d} \tilde{\Theta}(\ell + \mathbf{L}) \tilde{\Theta}(\ell)^* g_{\mathbf{L}}(\ell) d\ell$ where the weights g must satisfy the constraint $g_{\mathbf{L}}(\ell) f_{\mathbf{L}}(\ell) \geq 0$ and the normalizing constant $A_{\mathbf{L}}$ is given by $A_{\mathbf{L}}^{-1} \equiv \int g_{\mathbf{L}}(\ell) f_{\mathbf{L}}(\ell) d\ell$ to ensure unbiasedness. In [16] the optimal weights g are derived to be proportional to $f_{\mathbf{L}}(\ell)^*/[C_{\ell+\mathbf{L},\text{expt}}^{\Theta\Theta} C_{\ell,\text{expt}}^{\Theta\Theta}]$. Now, plugging in these optimal weights and using (5) and (6) one immediately gets

$$\begin{aligned} \hat{\phi}(\mathbf{L}) &\equiv (\mathbf{1}^\dagger \mathbf{N}_{\mathbf{L}}^{-1} \mathbf{1})^{-1} \mathbf{1}^\dagger \mathbf{N}_{\mathbf{L}}^{-1} \mathbf{v}_{\mathbf{L}} \\ &= A_{\mathbf{L}} \int_{\mathbb{R}^d} \tilde{\Theta}(\ell + \mathbf{L}) \tilde{\Theta}(\ell)^* g_{\mathbf{L}}(\ell) d\ell. \end{aligned}$$

Moreover, the prediction variance follows easily from standard regression theory which gives the following variance of $\hat{\phi}$

$$\begin{aligned} \langle \hat{\phi}(\mathbf{L}) \hat{\phi}(\mathbf{L}')^* \rangle - \phi(\mathbf{L}) \phi(\mathbf{L}')^* \\ = \begin{cases} (\mathbf{1}^\dagger \mathbf{N}_{\mathbf{L}}^{-1} \mathbf{1})^{-1}, & \text{if } \mathbf{L} = \mathbf{L}'; \\ 0, & \text{otherwise} \end{cases} \\ = \delta_{\mathbf{L}-\mathbf{L}'} 2A_{\mathbf{L}} \end{aligned}$$

where the last equality holds by (6).

III. SHRINKING THE QUADRATIC ESTIMATOR WHEN $C_L^{\phi\phi}$ IS KNOWN

The main reason the regression characterization of the quadratic estimator is useful is to easily see how one would relax the unbiasedness constraint when the true $C_L^{\phi\phi}$ is known. Indeed, a natural extension to the generalized regression estimate is called ridge regression which can be written

$$\tilde{\phi}_{\lambda_{\mathbf{L}}}(\mathbf{L}) \equiv (\mathbf{1}^\dagger \mathbf{N}_{\mathbf{L}}^{-1} \mathbf{1} + \lambda_{\mathbf{L}}^{-1})^{-1} \mathbf{1}^\dagger \mathbf{N}_{\mathbf{L}}^{-1} \mathbf{v}_{\mathbf{L}} \quad (7)$$

where $\lambda_{\mathbf{L}}$ is a ridge parameter which regularizes the intrinsic instability that arises when $\mathbf{1}^\dagger \mathbf{N}_{\mathbf{L}}^{-1} \mathbf{1}$ is small (i.e. when the variance of the re-construction is large). When $\phi(\mathbf{L})$ is a realization from a mean zero Gaussian random field and $\lambda_{\mathbf{L}}$ is set to the variance of $\phi(\mathbf{L})$ then the ridge regression estimate $\tilde{\phi}_{\lambda_{\mathbf{L}}}(\mathbf{L})$ can be interpreted as

the posterior expected value given the data $\mathbf{v}_{\mathbf{L}}$ under the assumption that both $\mathbf{e}_{\mathbf{L}}$ (defined in (3)) and $\phi(\mathbf{L})$ are uncorrelated, mean zero and jointly Gaussian. Returning to the original characterization of the quadratic estimate this leads to the following expression for $\tilde{\phi}_{\lambda_{\mathbf{L}}}(\mathbf{L})$:

$$\tilde{\phi}_{\lambda_{\mathbf{L}}}(\mathbf{L}) = \frac{1}{A_{\mathbf{L}}^{-1} + 2/C_L^{\phi\phi}} \int_{\mathbb{R}^d} \tilde{\Theta}(\ell + \mathbf{L}) \tilde{\Theta}(\ell)^* g_{\mathbf{L}}(\ell) d\ell$$

when setting $\lambda_{\mathbf{L}} = \delta_0 C_L^{\phi\phi}$.

Moreover, the posterior variance of $\phi(\mathbf{L})$, under this Bayesian interpretation, is simply $(\mathbf{1}^\dagger \mathbf{N}_{\mathbf{L}}^{-1} \mathbf{1} + \lambda_{\mathbf{L}}^{-1})^{-1} = \delta_0 / [(2A_{\mathbf{L}})^{-1} + (C_L^{\phi\phi})^{-1}]$.

The above Bayesian interpretation of $\tilde{\phi}_{\lambda_{\mathbf{L}}}(\mathbf{L})$ relies on the assumption that $\mathbf{v}_{\mathbf{L},k}$ is Gaussian. This is clearly violated since $\mathbf{v}_{\mathbf{L},k}$ is a product of two Gaussians. However, one can also interpret the ridge regression estimate $\tilde{\phi}_{\lambda_{\mathbf{L}}}(\mathbf{L})$ in the spirit of a Wiener filter of the quadratic estimator which makes the Gaussianity assumption less worrisome. In particular, if one treats the quadratic estimate $\hat{\phi}$ as data, the approximate unbiasedness of the quadratic estimator allows one to write

$$\hat{\phi}(\mathbf{L}) = \phi(\mathbf{L}) + \epsilon(\mathbf{L}) \quad (8)$$

where $\langle \epsilon(\mathbf{L}) \rangle \approx 0$ and $\langle \epsilon(\mathbf{L}) \epsilon(\mathbf{L}')^* \rangle = \delta_{\mathbf{L}-\mathbf{L}'} 2A_{\mathbf{L}}$. The supposition that ϵ is Gaussian becomes more believable since $\hat{\phi}$ is a weighted average of $\mathbf{v}_{\mathbf{L},k}$ which are approximately independent random variables (up to zero order in ϕ). If one now applies the Bayesian paradigm under the assumption that ϕ and ϵ are uncorrelated mean zero Gaussian random fields, then the posterior expected value of ϕ when observing $\hat{\phi}$ becomes

$$\tilde{\phi}_{\lambda_{\mathbf{L}}}(\mathbf{L}) = \frac{C_L^{\phi\phi}}{2A_{\mathbf{L}} + C_L^{\phi\phi}} \hat{\phi}(\mathbf{L}) \quad (9)$$

when setting $\lambda_{\mathbf{L}} = \delta_0 C_L^{\phi\phi}$.

This agrees with (7) and is clearly recognized as a Wiener filter of the quadratic estimate $\hat{\phi}$.

IV. ADAPTIVE SHRINKING WHEN $C_L^{\phi\phi}$ IS UNKNOWN

The previous section derived the optimal Wiener filter of the quadratic estimate when the spectral density $C_L^{\phi\phi}$ is known. In this section we study the scenario that $C_L^{\phi\phi}$ is not known. The main question is how to derive a shrinkage factor, as in the Wiener filter (9), without knowledge of $C_L^{\phi\phi}$. This is derived using a hierarchical Bayesian analysis which establishes that the shrinkage factor is a mixture over the possible values of the $C_L^{\phi\phi}$ supported by the data. In IV A we first discuss the full Bayesian analysis which requires a prior distribution on the unknown spectral density $C_L^{\phi\phi}$. To circumvent

computational difficulties with such an analysis we then recommend a non-informative generalized prior, in IV B, which gives rise our robust adaptive shrinkage estimator (1) presented in the introduction.

To align our notation with standard statistical theory it will be useful to concatenate the values of $\hat{\phi}(\mathbf{L})$, for different frequencies \mathbf{L} , into one data vector of length n , denoted $\hat{\phi}$. Define ϕ similarly as the vector of $\phi(\mathbf{L})$ values for the matching frequencies \mathbf{L} used to construct $\hat{\phi}$. We make the additional assumption that ϕ and $\hat{\phi}$ contain only unique elements up to complex conjugation (so there are no distinct coordinates with values w and z such that $z = w$ or $z = w^*$). Working with the vectors ϕ and $\hat{\phi}$, instead of the functions ϕ and $\hat{\phi}$, has the additional advantage that one can easily extend to the case where the adaptive shrinkage is done on separate annuli in Fourier space (which is discussed in Remark 1 at the end of Section IV B).

We work under the viewpoint, used to derived equation (9), that treats $\hat{\phi}$ as “data” which are then used to estimate ϕ . Translating equation (9) to our vector notation, the Wiener filter is given simply by matrix multiplication

$$\tilde{\phi}_\Lambda \equiv \Lambda(\Lambda + \Sigma)^{-1} \hat{\phi} \quad (10)$$

where the matrices Σ and Λ are defined by

$$\begin{aligned} \Sigma &\equiv \left(\delta_{\mathbf{L}_k - \mathbf{L}_j} 2A_{\mathbf{L}_k} \right)_{k,j=1}^n \\ \Lambda &\equiv \left(\delta_{\mathbf{L}_k - \mathbf{L}_j} C_{\mathbf{L}_k}^{\phi\phi} \right)_{k,j=1}^n. \end{aligned}$$

In addition, our notation allows us to clearly write the relationship between ϕ and $\hat{\phi}$ in a hierarchal Bayesian setting

$$\hat{\phi} | \phi \sim \mathcal{N}(\phi, \Sigma) \quad (11)$$

$$\phi | \Lambda \sim \mathcal{N}(0, \Lambda) \quad (12)$$

where $\phi | \Lambda \sim \mathcal{N}(0, \Lambda)$ means $\text{Re}(\phi)$ and $\text{Im}(\phi)$ are independent Gaussian random vectors with individual distributions given by $\text{Re}(\phi) \sim \mathcal{N}(0, \Lambda/2)$ and $\text{Im}(\phi) \sim \mathcal{N}(0, \Lambda/2)$.

A. The Bayes Solution

The Bayesian paradigm is the clearest way to understand how one adapts (10) when Λ is unknown. Indeed, if one is willing to model the uncertainty in Λ (equivalently in $C_L^{\phi\phi}$) using a prior probability density $\mathcal{P}(\Lambda)$, then Bayes theorem in conjunction with (11) and (12) gives a posterior density $\mathcal{P}(\phi, \Lambda | \hat{\phi}) \propto \mathcal{P}(\hat{\phi} | \phi) \mathcal{P}(\phi | \Lambda) \mathcal{P}(\Lambda)$. The posterior density $\mathcal{P}(\phi, \Lambda | \hat{\phi})$ quantifies the joint uncertainty in the unobserved ϕ and Λ when observing the data $\hat{\phi}$. Notice that one can marginalize out Λ to obtain a posterior on ϕ exclusively, $\mathcal{P}(\phi | \hat{\phi}) =$

$\int \mathcal{P}(\phi, \Lambda | \hat{\phi}) d\Lambda = \int \mathcal{P}(\phi | \Lambda, \hat{\phi}) \mathcal{P}(\Lambda | \hat{\phi}) d\Lambda$ where $d\Lambda$ corresponds to coordinate-wise area element. Now the posterior expected value of ϕ given the data $\hat{\phi}$ can be computed as

$$\begin{aligned} \int \phi \mathcal{P}(\phi | \hat{\phi}) d\phi &= \int \underbrace{\left[\int \phi \mathcal{P}(\phi | \Lambda, \hat{\phi}) d\phi \right]}_{\text{Wiener filter (10)}} \mathcal{P}(\Lambda | \hat{\phi}) d\Lambda \\ &= \int \tilde{\phi}_\Lambda \mathcal{P}(\Lambda | \hat{\phi}) d\Lambda. \end{aligned} \quad (13)$$

The advantage of (13) is that it clearly exposes how to handle the situation when Λ is unknown: average the Wiener filter $\tilde{\phi}_\Lambda$ over different possibilities for Λ supported by the data (through the posterior $\mathcal{P}(\Lambda | \hat{\phi})$). In fact, since $\tilde{\phi}_\Lambda$ depends on Λ only through a multiplicative factor, (13) simplifies to

$$\int \tilde{\phi}_\Lambda \mathcal{P}(\Lambda | \hat{\phi}) d\Lambda = \underbrace{\left[\int \Lambda(\Lambda + \Sigma)^{-1} \mathcal{P}(\Lambda | \hat{\phi}) d\Lambda \right]}_{\text{posterior expected shrinkage factor}} \hat{\phi} \quad (14)$$

Therefore, to account for uncertainty in Λ when estimating ϕ simply replace the shrinkage factor $\Lambda(\Lambda + \Sigma)^{-1}$ in (10) with it's expected value under the posterior distribution on Λ , $\mathcal{P}(\Lambda | \hat{\phi})$.

B. Robust generalized Bayesian adaptive shrinkage

The Bayesian analysis, while being a complete probabilistic combination of the data $\hat{\phi}$ and the prior knowledge for Λ , can be time consuming for two reasons. First, one needs to translate the knowledge in Λ to a prior distribution $\mathcal{P}(\Lambda)$. Secondly, getting posterior samples from $\mathcal{P}(\Lambda | \hat{\phi})$ often requires advanced Monte Carlo techniques (similar to those found in [8, 37] for example). For these reasons we present a simplified non-informative generalized prior on Λ , originally developed in [3, 36], for quick exploratory analysis. This prior requires the user to specify a fiducial spectral density model, denoted $C_{L, \text{fid}}^{\phi\phi}$, and is designed to be both robust against alternative spectral truths and computationally simple.

To specify the prior start by defining the following matrix based on the fiducial spectral density

$$\Lambda_{\text{fid}} \equiv \left(\delta_{\mathbf{L}_k - \mathbf{L}_j} C_{\mathbf{L}_k, \text{fid}}^{\phi\phi} \right)_{k,j=1}^n.$$

The non-informative generalized prior for Λ is given by

$$\Lambda \sim \xi(\Sigma + \Lambda_{\text{fid}}) - \Sigma \quad (15)$$

$$\xi \text{ has generalized density } \propto \xi^{-1} \text{ on } (\rho, \infty) \quad (16)$$

where ρ is the largest eigenvalue of $\Sigma(\Sigma + \Lambda_{\text{fid}})^{-1}$ (which ensures that $\xi(\Sigma + \Lambda_{\text{fid}}) - \Sigma$ is always positive definite). Since the support of ξ contains 1, the prior includes the

fiducial model Λ_{fid} as a possible truth. The generalized density ξ^{-1} is an improper prior (it integrates to infinity) which can be derived as the well known Jeffereys' non-informative prior within the class of distributions $\Lambda \sim \xi(\Sigma + \Lambda_{\text{fid}}) - \Sigma$ for the hierarchal parameter ξ . The theoretical properties of this and similar priors have been extensively studied in the statistical literature (see [3, 4, 36]) and has been shown to yield estimators with desirable statistical properties.

One of the advantages of this prior is that marginalizing over ϕ and Λ collapses (11), (12) and (15) to

$$\hat{\phi}|\xi \sim \mathcal{N}(0, \xi(\Sigma + \Lambda_{\text{fid}})). \quad (17)$$

Now it is easy to obtain the posterior distribution on ξ . In particular, let $\mathcal{P}(\hat{\phi}|\xi, \Lambda_{\text{fid}}, \Sigma)$ denote the marginal density of $\hat{\phi}$ in (17) so that

$$\mathcal{P}(\hat{\phi}|\xi, \Lambda_{\text{fid}}, \Sigma) \propto \frac{1}{\xi^n |\Sigma + \Lambda_{\text{fid}}|} \exp\left(-\|\hat{\phi}\|_{\text{fid}}^2 / \xi\right)$$

where $\|\hat{\phi}\|_{\text{fid}}^2$ is defined by

$$\begin{aligned} \|\hat{\phi}\|_{\text{fid}}^2 &\equiv \left\| (\Sigma + \Lambda_{\text{fid}})^{-1/2} \text{Re}(\hat{\phi}) \right\|^2 \\ &\quad + \left\| (\Sigma + \Lambda_{\text{fid}})^{-1/2} \text{Im}(\hat{\phi}) \right\|^2. \end{aligned}$$

Note, the number of elements in $\hat{\phi}$ is n (so that the vector $(\text{Re}(\hat{\phi})^\dagger, \text{Im}(\hat{\phi})^\dagger)^\dagger$ has $2n$ entries, for example). Therefore a formal application of Bayes theorem for ξ under the model (17) gives

$$\begin{aligned} \mathcal{P}(\xi|\hat{\phi}) &\propto \mathcal{P}(\hat{\phi}|\xi, \Lambda_{\text{fid}}, \Sigma) \mathcal{P}(\xi) \\ &\propto \frac{1}{\xi^{n+1}} \exp\left(-\|\hat{\phi}\|_{\text{fid}}^2 / \xi\right) \end{aligned} \quad (18)$$

on (ρ, ∞) where $\mathcal{P}(\xi) \propto \xi^{-1}$ denotes the prior for ξ . This shows that the posterior $\mathcal{P}(\xi|\hat{\phi})$ has a truncated inverse gamma distribution, which can easily be sampled from, as in Algorithm 1 below.

Algorithm 1: Sample from $\mathcal{P}(\xi|\hat{\phi})$

- 1: Simulate a Gamma random variable ζ with density proportional to $\zeta^{n-1} \exp(-\zeta \|\hat{\phi}\|_{\text{fid}}^2)$
 - 2: **if** $\zeta^{-1}(\Sigma + \Lambda_{\text{fid}}) - \Sigma$ is positive definite **then** go to step 3, **else** go back to step 1
 - 3: **return** $\xi \leftarrow \zeta^{-1}$
-

The posterior samples from $\mathcal{P}(\xi|\hat{\phi})$ do not have a direct physical interpretation. However, Algorithm 2 shows how sampling from $\mathcal{P}(\xi|\hat{\phi})$ allows easy sampling from $\mathcal{P}(\phi|\hat{\phi})$. Actually, it is not immediately obvious that Algorithm 2 gives samples from $\mathcal{P}(\phi|\hat{\phi})$ since the prior $\mathcal{P}(\xi)$ is improper. However, a careful application of Fubini shows, indeed, Algorithm 2 samples $\mathcal{P}(\phi|\hat{\phi})$. In addition, the

posterior expected shrinkage factor in (14) can be computed as follows

$$\begin{aligned} &\int \Lambda(\Sigma + \Lambda)^{-1} \mathcal{P}(\Lambda|\hat{\phi}) \\ &= \int [I - \xi^{-1} \Sigma(\Sigma + \Lambda_{\text{fid}})^{-1}] \mathcal{P}(\xi|\hat{\phi}) d\xi \\ &= I - F^{\hat{\phi}} \Sigma(\Sigma + \Lambda_{\text{fid}})^{-1} \end{aligned} \quad (19)$$

where $F^{\hat{\phi}} \equiv \int \xi^{-1} \mathcal{P}(\xi|\hat{\phi})$ and I denotes the $n \times n$ identity matrix. Keeping track of the normalization factor in (18), one obtains the following analytic expression for $F^{\hat{\phi}}$

$$F^{\hat{\phi}} = \frac{n-1}{\|\hat{\phi}\|_{\text{fid}}^2} \frac{P(n, \|\hat{\phi}\|_{\text{fid}}^2 / \rho)}{P(n-1, \|\hat{\phi}\|_{\text{fid}}^2 / \rho)} \quad (20)$$

where $P(a, x)$ is the normalized incomplete gamma function given by $P(a, x) \equiv \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt$. Combining equations (19) and (14) one obtains the following adaptive shrinkage estimate of ϕ given $\hat{\phi}$

$$\tilde{\phi} \equiv \int \phi \mathcal{P}(\phi|\hat{\phi}) = [I - F^{\hat{\phi}} \Sigma(\Sigma + \Lambda_{\text{fid}})^{-1}] \hat{\phi} \quad (21)$$

which is recognized as the matrix form of (1).

Remark 1: In our implementation of (21) we partition the Fourier frequencies \mathbf{L} into concentric annuli around the origin and construct the posteriors $\mathcal{P}(\xi|\hat{\phi})$ and $\mathcal{P}(\phi|\hat{\phi})$ separately on each annuli. In this way we obtain distinct shrinkage factor adjustments $F^{\hat{\phi}}$ for each annuli, hence the dependence of $F^{\hat{\phi}}$ on L , written $F_L^{\hat{\phi}}$ in (1). This essentially adds flexibility by allowing independent priors $\mathcal{P}(\xi)$ for each annuli. In fact, if $2A_{\mathbf{L}}$ changes drastically within an annuli, one may further partition with the goal of obtaining partitions within which the values of $2A_{\mathbf{L}}$ are similar.

Remark 2: Although, one has access to an analytic expression for the $F^{\hat{\phi}}$ given in (20), this formula becomes numerically unstable when either $\|\hat{\phi}\|_{\text{fid}}^2$ is small or n large. Therefore we recommend approximating $F_L^{\hat{\phi}} = \int \xi^{-1} \mathcal{P}(\xi|\hat{\phi}) d\xi$ by averaging samples of ξ^{-1} obtained from Algorithm 1.

Algorithm 2: Sample from $\mathcal{P}(\phi|\hat{\phi})$

- 1: Simulate ξ from Algorithm 1
- 2: Set $\Lambda \leftarrow \xi(\Sigma + \Lambda_{\text{fid}}) - \Sigma$ and simulate

$$\phi \sim \mathcal{N}\left(\Lambda(\Sigma + \Lambda)^{-1} \hat{\phi}, [\Sigma^{-1} + \Lambda^{-1}]^{-1}\right)$$

- 3: **return** ϕ
-

Remark 3: To connect the Bayes estimator (21) with alternative non-Bayesian estimates of ϕ , it is instructive to consider the limit as $n \rightarrow \infty$. The cleanest connection

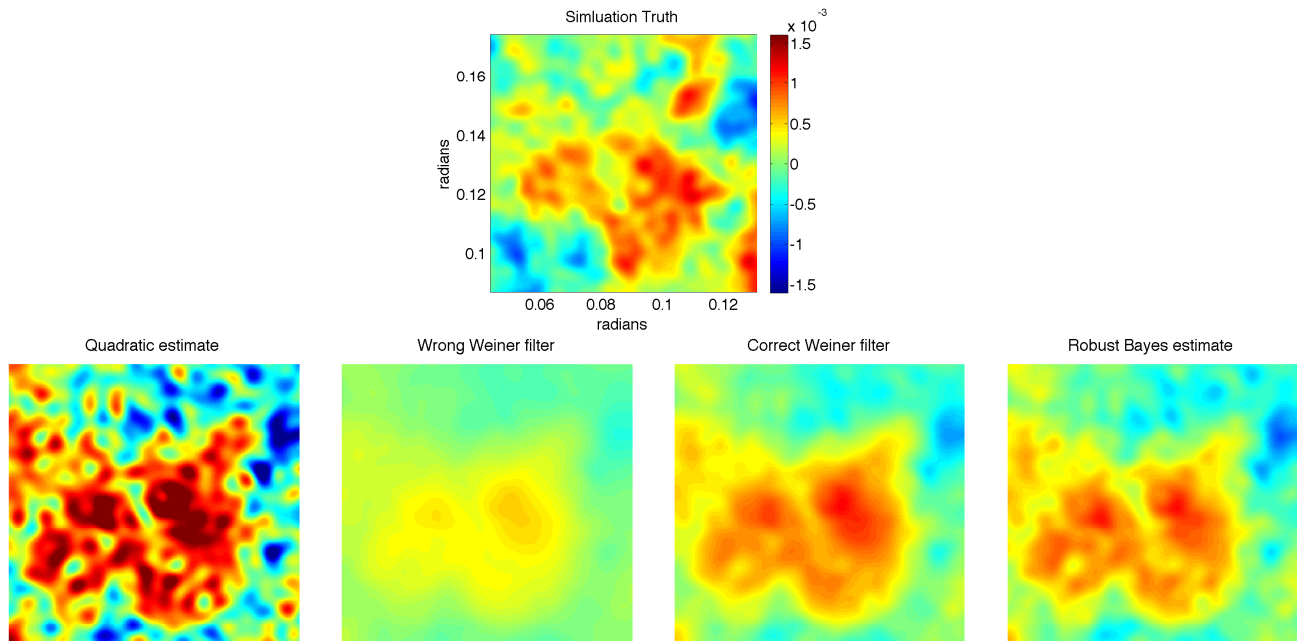


FIG. 1. **Example 1:** Pixel space estimates (bottom row) of the filtered lensing potential $|\mathbf{L}|\phi(\mathbf{L})\mathbf{1}_{|\mathbf{L}|\leq 1000}$ (top). The lensed simulation is done with arcmin pixels, a beam FWHM set to 1 arcmin and additive white noise with a standard deviation of $25 \mu K\text{-arcmin}$. The quadratic estimate is shown bottom left and the robust Bayes estimate is shown bottom right using an input fiducial spectral density $C_{L,\text{fid}}^{\phi\phi}$ set 4 times smaller than the simulation truth spectral density $C_L^{\phi\phi}$. The bottom middle two plots show the Wiener filter of the quadratic estimate given by (9) based on $C_{L,\text{fid}}^{\phi\phi}$ (middle left) and the true but unknown spectral density $C_L^{\phi\phi}$ (middle right).

occurs when the fiducial model $C_{L,\text{fid}}^{\phi\phi}$ is set to 0 (so that $\Lambda_{\text{fid}} = 0$) which postulates no lensing in the observed CMB. Making the additional assumption $\Sigma = 2\sigma^2 I$ the estimate (21) simplifies to $\tilde{\phi} = [1 - F^{\hat{\phi}}]\hat{\phi}$. In Appendix B we show

$$[1 - F^{\hat{\phi}}] = \left[1 - \frac{(2n-2)\sigma^2}{\hat{\phi}^\dagger \hat{\phi}^*}\right]^+ + o(1) \quad (22)$$

where $o(1) \rightarrow 0$ uniformly in $\hat{\phi}$ as $n \rightarrow \infty$ and $x^+ \equiv \max\{0, x\}$. The right hand side of (22) is the shrinkage factor used in the famous James-Stein shrinkage estimator [18]. This follows since the real and imaginary parts of $\hat{\phi}$ are modeled as $2n$ independent Gaussian random variables, each with variance σ^2 . In fact, for any fixed number $\epsilon > 0$, the supremum of $|o(1)|$ over the region $\frac{\hat{\phi}^\dagger \hat{\phi}^*}{2n-2} \geq \sigma^2 + \epsilon$ converges to zero exponentially fast as $n \rightarrow \infty$. Therefore if one changes $2n-2$ in the numerator of (22) to, say $2n$, this exponential convergence fails to hold. Even more is true: under the same assumptions on Λ_{fid} and Σ , the results of [2] show that $\tilde{\phi}$ is a min-max and admissible estimate of ϕ with respect to the quadratic loss when $n \geq 2$. Similar results for (21) and other Bayes estimates have been derived in the statistical literature (see [4, 25] for a review of the literature).

V. SIMULATIONS

We present three simulation examples which give an overview of some of the advantages provided by the robust Bayesian procedure developed in Section IV. The first example demonstrates robustness and compares with the optimal Wiener filter. The second compares point-wise error quantification with the quadratic estimator. The final example explores spectral power estimation and demonstrates the robust Bayesian method can generate more accurate detection levels than the quadratic version based on spectral density estimation.

Example 1

In this example we consider the problem of imaging estimated over-densities (in pixel space) of the filtered potential. We focus on the filter $|\mathbf{L}|\phi(\mathbf{L})\mathbf{1}_{|\mathbf{L}|\leq 1000}$ where $\mathbf{1}_{|\mathbf{L}|\leq 1000}$ is the top hat indicator with radius 1000. We remove the spectral power beyond 1000 for two reasons. First, the high frequency power must be removed from the quadratic estimate or else the noise completely dominates the image. Secondly the, so called, $N_L^{(1)}$ bias starts to significantly contaminate the results beyond 1000 (see [10, 22]).

The bottom row of images in FIG. 1 show four dif-

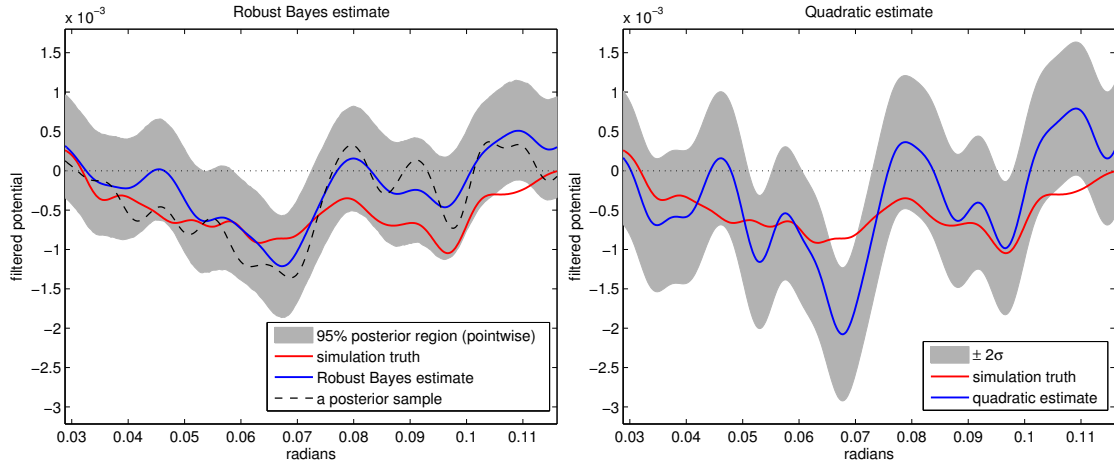


FIG. 2. **Example 2:** Pixel space line transect plots of the robust Bayes estimate (left, blue) and quadratic estimate (right, blue) of the filtered lensing potential $|\mathbf{L}|\phi(\mathbf{L})\mathbf{1}_{|\mathbf{L}|\leq 1000}$ plotted in red for both left and right diagrams. The shaded region corresponds to 2σ error bars on the right and a point-wise 95% posterior region on the left. The lensed simulation is done with arcmin pixels, a beam FWHM set to 4 arcmin and additive white noise with a standard deviation of $5\mu K$ -arcmin. The input fiducial spectral density $C_{L,\text{fid}}^{\phi\phi}$ used for the robust Bayes estimate is set 4 times larger than the simulation true spectral density.

ferent estimates of the filtered lensing potential in pixel space based on a lensed CMB simulation in a $17^\circ \times 17^\circ$ patch of periodic flat sky observed on arcmin pixels. The lensed CMB simulations are done using a Gaussian beam FWHM of 1 arcmin and additive white noise with standard deviation $25\mu K$ -arcmin (further simulation details can be found in Appendix A). The top plot shows the simulation truth, zoomed in on an over dense region. The bottom left shows the corresponding region for the quadratic estimate and the bottom right shows the robust Bayesian estimate given in (1) where the fiducial spectral density $C_{L,\text{fid}}^{\phi\phi}$ is 4 times smaller than simulation truth spectral density $C_L^{\phi\phi}$. The middle two plots show Wiener filters of the quadratic estimate given by (9) based on the true, but unknown spectral density $C_L^{\phi\phi}$ (middle right) and on the same fiducial spectral density $C_{L,\text{fid}}^{\phi\phi}$ (middle left) used to generate the robust Bayes estimate.

There are three things to notice here. First, the robust Bayes estimate successfully shrinks the noisy high frequency terms in the quadratic estimate (i.e. far left is noisier than than far right). Secondly, if one simply used the Wiener filter (9) based on the wrong fiducial model $C_{L,\text{fid}}^{\phi\phi}$ one would seriously over shrink the quadratic estimate (i.e. middle left over shrinks). Three, the adaptivity factor $F_L^{\hat{\phi}}$ is adjusting the robust Bayes estimate to behave more like the optimal Wiener filter (9) when one has access to the true spectral density $C_L^{\phi\phi}$ (i.e. middle right and far right look similar). This demonstrates some robustness to mis-specification of the fiducial model.

Example 2

The second simulation compares point-wise error quantification for estimating the same filtered potential $|\mathbf{L}|\phi(\mathbf{L})\mathbf{1}_{|\mathbf{L}|\leq 1000}$ as in last example. To contrast with the last example, we use a different fiducial spectral density: one that is 4 times too large rather than 4 times too small. We also change the beam FWHM to 4 arcmin and additive noise standard deviation to $5\mu K$ -arcmin (the pixel size and sky coverage is the same as the last example). FIG. 2 shows line transect plots of the filtered simulation truth, plotted in red for both left and right diagrams. The right diagram shows the quadratic estimate (blue) with 2σ error regions shaded grey. The 2σ region is computed using the Fourier space variance approximation $2|\mathbf{L}|^2 A_L \mathbf{1}_{|\mathbf{L}|\leq 1000}$ for the filtered quadratic estimate. The left diagram shows the robust Bayesian estimate (blue) with the shading region denoting a 95% the posterior region (point-wise) which is determined from simulation. We also included a plot of one posterior posterior sample (dashed). The point-wise error bars corresponding to the Bayes estimate are somewhat smaller than those from the quadratic estimator. However, the main feature of these plots is that, even though the fiducial model is 4 times too large, the Bayes estimate successfully shrinks the noisy high frequency terms in the quadratic estimator. One can also see the advantage of having posterior realizations of possible truths supported by the data (dashed) which allow joint quantification of uncertainty rather than simple point-wise mean and standard error bars provided by the quadratic estimate.

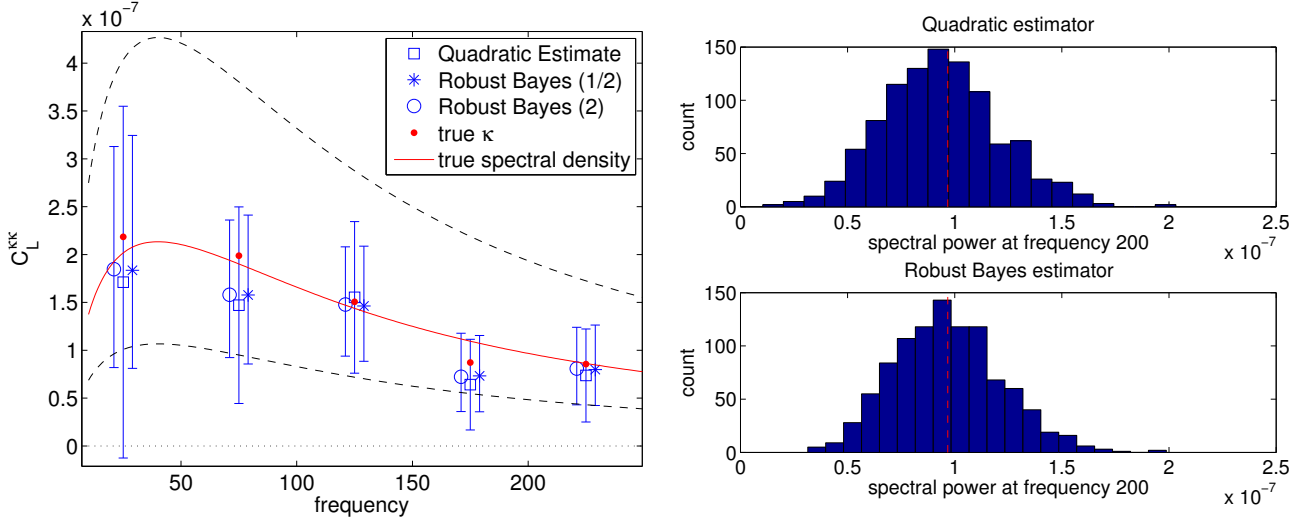


FIG. 3. **Example 3:** The left plot illustrates spectral power estimation for the convergence $\kappa(\mathbf{L})$ using the quadratic estimate (squares) and the robust Bayesian method (circles and stars) for two different input fiducial models (dashed lines). The error bars for the Bayesian estimates correspond to 2.5 and 97.5 posterior percentiles, whereas the quadratic estimate is displayed with 2σ error bars (see Section V for simulation details). The histograms, at right, show the quadratic estimates and the Bayes estimates of the spectral power in $\kappa(\mathbf{L})$ at wavelength $L = 200$ based on 1000 independent lensing simulations. The dashed red vertical lines in the histogram plots designate the value of the true spectral density $C_{200}^{\kappa\kappa}$. See Section V for further details.

Example 3

In this example we explore spectral power detection and demonstrate that the robust Bayesian method can generate more accurate detection levels than the quadratic version based on spectral density estimation. CMB gravitational lensing surveys often focus on detecting non-zero spectral density power in the so called convergence field $\kappa(\mathbf{x}) \equiv -\nabla^2\phi(\mathbf{x})/2$ (using the notation found in [38]) which is a tracer of projected mass. Typically, the quadratic estimate is used to estimate the spectral density $C_L^{\kappa\kappa}$ which, along with standard error bars, are then used to report σ -detection levels. However, when estimating the spectral density there is always unavoidable cosmic variance, which limits the σ -detection level achievable. To illustrate this fact, consider the case that one is able to observe the $\kappa(\mathbf{x})$ field directly, with no noise. In this case there is no uncertainty in lensing detection. However, there is still cosmic variance uncertainty for estimating the spectral density $C_L^{\kappa\kappa}$. We explore this issue by taking advantage of the robust Bayes method which easily produces estimates of the spectral power in κ directly, instead of through $C_L^{\kappa\kappa}$. We demonstrate better σ -detection levels for gravitational lensing than would be obtained through quadratic estimates of $C_L^{\kappa\kappa}$.

The parameter of interest, in this case, is the spectral power in κ . Under the isotropic assumption on κ it is natural to radially average the spectral power over concentric annuli about the origin. We only consider annuli which are cut in half since $\kappa(-\mathbf{L}) = \kappa(\mathbf{L})^*$, because $\kappa(\mathbf{x})$ is a real random field, which implies redundant informa-

tion on the other half of the annuli. In particular, for each half annuli \mathcal{A} , we want to estimate the quantity

$$\frac{\Delta\mathbf{L}}{\#\mathcal{A}} \sum_{\mathbf{L} \in \mathcal{A}} |\kappa(\mathbf{L})|^2 \quad (23)$$

where $\Delta\mathbf{L}$ denotes the area of the grid spacing in Fourier space arising from finite sky observations, and $\#\mathcal{A}$ denotes the number of observed Fourier frequencies in \mathcal{A} . Notice the factor $\Delta\mathbf{L}$ makes (23) an unbiased estimate of $C_L^{\kappa\kappa}$ when the annuli radii and $\Delta\mathbf{L}$ are infinitesimally small (this follows by the equality $\langle \kappa(\mathbf{L})\kappa(\mathbf{L}')^* \rangle_\phi = \delta_{\mathbf{L}-\mathbf{L}'} C_L^{\kappa\kappa}$ where $\langle \cdot \rangle_\phi$ denotes expectation with respect to ϕ).

The Bayesian estimate for (23) is exceedingly easy to construct: simply average the quantity (23) computed on each posterior sample ϕ obtained from Algorithm 2. To compare with the quadratic estimate of the spectral density $C_L^{\kappa\kappa}$, which is typically used for detection, we use the following unbiased (up to leading order) quadratic estimate of $C_L^{\kappa\kappa}$

$$\widehat{C}_L^{\kappa\kappa} \equiv \frac{\Delta\mathbf{L}}{\#\mathcal{A}} \sum_{\mathbf{L} \in \mathcal{A}} |\hat{\kappa}(\mathbf{L})|^2 - \frac{1}{\#\mathcal{A}} \sum_{\mathbf{L} \in \mathcal{A}} |\mathbf{L}|^4 A_L/2 \quad (24)$$

where $\hat{\kappa}(\mathbf{x}) \equiv -\nabla^2\hat{\phi}(\mathbf{x})/2$ and $\hat{\phi}$ is the quadratic estimate. Under the Gaussian approximation one gets the following variance for $\widehat{C}_L^{\kappa\kappa}$

$$\text{var } \widehat{C}_L^{\kappa\kappa} = \frac{1}{(\#\mathcal{A})^2} \sum_{\mathbf{L} \in \mathcal{A}} (C_L^{\kappa\kappa} + |\mathbf{L}|^4 A_L/2)^2 \quad (25)$$

This approximation agrees with [22] and [10] (in the curved sky) when $2A_L$ is rotationally symmetric (the

missing factor of 2 found in [22] and [10] appears since our \mathcal{A} is only half an annulus). Notice that to compute the variance of $\widehat{C}_L^{\kappa\kappa}$ one must know the unknown spectral density $C_L^{\kappa\kappa}$. To approximate $\text{var } \widehat{C}_L^{\kappa\kappa}$ we therefore replace $C_L^{\kappa\kappa}$ in the right hand side of equation (25) with the estimate $\widehat{C}_L^{\kappa\kappa}$.

Our first comparison of the Bayes and quadratic estimates of (23) is shown in the left plot of FIG. 3 and is based on one realization of a lensed CMB field ($17^\circ \times 17^\circ$ periodic sky, arcmin pixels, beam FWHM = 1 arcmin, white $25 \mu K$ -arcmin noise) with spectral density $C_L^{\kappa\kappa}$ shown in red. We computed two Bayes estimates: one which uses a fiducial model $C_{L,\text{fid}}^{\phi\phi}$ which is 2 times too large (estimates are denoted with circles, fiducial model with dashed line), and one using $C_{L,\text{fid}}^{\phi\phi}$ which is 2 times too large (estimates are denoted with stars, fiducial model with dashed line). The estimates are based on the first 5 concentric half annuli (width = 50) about the origin. The attached error bars correspond to 2.5th and 97.5th posterior percentiles. The squares show the quadratic estimate (24) with 2σ error bars based on (25). There are two points here. First, the Bayesian error bars are smaller and more naturally handle the positivity constraint. Secondly, the Bayesian regions are relatively robust with respect to misspecification of $C_{L,\text{fid}}^{\phi\phi}$ (i.e. the error bars attached to the circles and the stars are about the same size).

For our second comparison we check the posterior coverage probabilities, the quadratic estimate coverage probabilities and the relative sizes of the quadratic and Bayes error bars. We simulated 1000 independent lensed CMB realizations (with the same simulation configuration as above) using independent realizations of the unlensed CMB, the noise and the lensing potential for each simulation. The right hand plot of FIG. 3 shows the histograms of the corresponding quadratic and Bayes estimates of (23) at frequency $|\mathbf{L}| = 200$ (with half annuli width set to 50). The Bayes estimates were generated using a fiducial spectral density which is 2 times smaller than the input spectral density used for simulation. The dashed red vertical lines designate the value of the true spectral density $C_{200}^{\kappa\kappa}$.

The two histograms in FIG. 3 are similar with a slight variance and bias reduction in the Bayes estimates (15% smaller variance and 50% smaller bias). However, the main difference can be found in the associated error bars. For the Bayes estimate we use a posterior region based on 2.5th and 97.5th posterior percentiles for estimating (23), not $C_{200}^{\kappa\kappa}$. Surprisingly, exactly 950 of the 1000 posterior regions contained the true value of (23), which is different in each simulation. Conversely, 934 out of the 1000 2σ -error bars associated with the quadratic estimate covered the true value of $C_{200}^{\kappa\kappa}$. One would expect typical Monte Carlo fluctuations of about ± 7 (if true coverage was 95%) so this gives moderate evidence that the quadratic estimate error bars slightly undercover the truth. Moreover, the width of the posterior regions are 15% smaller than

the corresponding error bars for the quadratic estimate. This is due to the fact that the Bayes estimate does not contain cosmic variance uncertainty between (23) and $C_{200}^{\kappa\kappa}$.

It should be noted that it is possible to adjust the variance calculation (25) to represent the uncertainty for estimating (23), rather than the uncertainty for $C_L^{\kappa\kappa}$. This is somewhat beside the point, however, that the Bayesian estimates do this naturally no matter what functional one is estimating. There still needs to be a verification process that the Bayesian method behaves appropriately, however, for any rigorous scientific application. This most likely would include some type of simulation study. It may also include, as is done in Remark 3 of Section IV B, a derivation of the frequentest properties of the Bayesian estimates.

VI. DISCUSSION

The unbiasedness constraint in the quadratic estimator forces large variability in the estimated gravitational potential, especially at high frequency. This can present difficulties for exploratory data analysis, mapping the lensing potential and estimating nonlinear functionals of the gravitational potential, for example. In this paper we study the potential advantages obtained by relaxing the unbiasedness constraint in the quadratic estimator. To accomplish this we derived a regression characterization of the quadratic estimate which then allows one to clearly see how bias can be introduced for a reduction of variance: through ridge regression and Bayesian techniques. The Bayesian framework is especially cogent, essentially treating the quadratic estimate as data while incorporating prior information on $C_L^{\phi\phi}$. The resulting estimate is an adaptive Wiener filter adjustment to the raw quadratic estimate—shrinking frequencies with small SNR and retaining those with high SNR. As an alternative to a full Bayesian analysis, which can be somewhat demanding, we present a non-informative prior which not only leads to estimates with desirable frequentist properties (such as an asymptotic James-Stein shrinkage behavior) but also yields a posterior distribution which is easy to simulate without resorting to Monte Carlo techniques. The non-informative prior requires the user to input a fiducial model, but is designed to be robust to misspecification of this input model.

One clear advantage of the Bayesian analysis is the availability of posterior samples to construct estimates and uncertainty quantification for any non-linear functional of the gravitational potential, including spectral density estimation. Indeed, rather than using the shrinkage formula (1) directly (it becomes numerically unstable) we recommend averaging samples from the posterior obtained from algorithms 1 and 2 given in Section IV B. In Section V we explored the advantages gained by having easily obtained posterior samples: estimating and quantifying uncertainty in functionals of the gravitational po-

tential and joint quantification of uncertainty.

A drawback of the above Bayesian analysis—indeed of the quadratic estimate itself—is the bias obtained from the Taylor truncation on the lensed CMB used to derive the quadratic estimate. The spectral properties of this bias is relatively well understood [10, 22] when marginalizing over the randomness inherent in the large scale structure. However, it is unclear how this bias effects other features of the estimate of ϕ . Therefore, when applying the Bayesian methods described in the paper, we recommend avoiding the frequencies which are shown to be contaminated by biases. An alternative to avoiding frequencies is adjust $2A_L$ to include the so called $N_L^{(1)}$ and $N_L^{(2)}$ biases. However, until we have a good understanding of the sensitivity of $N_L^{(1)}$ and $N_L^{(2)}$ to a fiducial model this remains an unexplored possibility.

Appendix A: Simulation Details

The fiducial cosmology used in our simulations is based on a flat, power law Λ CDM cosmological model, with baryon density $\Omega_b = 0.044$; cold dark matter density $\Omega_{\text{cdm}} = 0.21$; cosmological constant density $\Omega_\Lambda = 0.74$; Hubble parameter $h = 0.71$ in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$; primordial scalar fluctuation amplitude $A_s(k = 0.002 \text{ Mpc}^{-1}) = 2.45 \times 10^{-9}$; scalar spectral index $n_s(k = 0.002 \text{ Mpc}^{-1}) = 0.96$; primordial helium abundance $Y_P = 0.24$; and reionization optical depth $\tau_r = 0.088$. The CAMB code is used to generate the theoretical power spectra [26].

To construct the lensed CMB simulation used in this paper we first generate a high resolution simulation of $\Theta(\mathbf{x})$ and the gravitational potential $\phi(\mathbf{x})$ on a periodic $17^\circ \times 17^\circ$ patch of the flat sky with 0.25 arcmin pixels. The lensing operation is performed by taking the numerical gradient of ϕ , then using linear interpolation to obtain the lensed field $\Theta(\mathbf{x} + \nabla\phi(\mathbf{x}))$. We down-sample the lensed field, every 4th pixel to obtain the desired arcmin pixel resolution for the simulation output. A Gaussian beam is then applied in Fourier space using FFT of the lensed fields. Finally white noise is added in pixel space.

Appendix B: Derivation of equation (22)

Start by letting $s \equiv \|\hat{\phi}\|_{\text{fid}}^2 / (n-1)$ which simplifies to $\frac{\hat{\phi}^\dagger \hat{\phi}^*}{\sigma^2(2n-2)}$ when $\Lambda_{\text{fid}} = 0$ and $\Sigma = 2\sigma^2 I$. Notice that ‘s’ behaves like a ratio between the observed signal power and the nominal noise level. To make the following calculations more readable we also let $m = n-1$. To establish (22) it will be sufficient to show

$$1 - \frac{1}{s} \frac{P(m+1, ms)}{P(m, ms)} = \left[1 - \frac{1}{s}\right]^+ + o(1) \quad (\text{B1})$$

where $o(1)$ is a function of both m and s such that

$$\sup_{s>0} |o(1)| \rightarrow 0 \quad \text{as } m \rightarrow \infty. \quad (\text{B2})$$

Notice that $1 - \frac{1}{s} \frac{P(m+1, ms)}{P(m, ms)}$ is increasing in s . This was shown in Lemma 2.1.1(vii) of [3] where it is noted that the density for the random variable ξ^{-1} has a decreasing monotone likelihood ratio in s . This is sufficient for stochastic domination (see Lemma 3.4.2 in [24], for example) which implies the expected value $\langle \xi^{-1} | \hat{\phi} \rangle \equiv \int \xi^{-1} \mathcal{P}(\xi | \hat{\phi}) = \frac{1}{s} \frac{P(m+1, ms)}{P(m, ms)}$ goes down when s goes up. Therefore the supremum $\sup_{0 < s \leq 1} |o(1)|$ is attained at $s = 1$. This follows by the fact that $1 - \frac{1}{s} \frac{P(m+1, ms)}{P(m, ms)}$ is increasing and positive (because it equals $1 - \langle \xi^{-1} | \hat{\phi} \rangle$ and ξ^{-1} is supported in $(0, 1]$). Therefore to prove (B2) it will be sufficient to establish $\sup_{s \geq 1} |o(1)| \rightarrow 0$ since

$$\sup_{0 < s \leq 1} |o(1)| \leq \sup_{s \geq 1} |o(1)|.$$

To study the case $s \geq 1$, integration by parts gives

$$\begin{aligned} 1 - \frac{1}{s} \frac{P(m+1, ms)}{P(m, ms)} &= 1 - \frac{1}{s} \frac{P(m, ms) - \frac{(ms)^m e^{-ms}}{m!}}{P(m, ms)} \\ &= 1 - \frac{1}{s} + \frac{m^m}{m!} \frac{(se^{-s})^m}{sP(m, ms)}. \end{aligned}$$

Therefore

$$\begin{aligned} \sup_{s \geq 1} |o(1)| &= \frac{(m/e)^m}{m!} \sup_{s \geq 1} \frac{(se^{1-s})^m}{sP(m, ms)} \\ &\leq \frac{(m/e)^m}{m!P(m, m)} \quad (\text{B3}) \end{aligned}$$

$$\sim \frac{2}{\sqrt{2\pi m}} \rightarrow 0 \quad (\text{B4})$$

Line (B3) follows since $se^{1-s} \leq 1$ (with equality only when $s = 1$) and the fact that $P(m, x)$ is an integral of a positive function over the interval $(0, x]$. Line (B4) follows by Stirling’s approximation and the fact that $P(m, m) = \text{Pr}[\sum_{k=1}^m X_k \leq m]$ where X_k are independent Gamma random variables with mean and variance 1. The central limit theorem then says $P(m, m) = \text{Pr}[\sqrt{m} \sum_{k=1}^m (X_k - 1) \leq 0] \rightarrow \text{Pr}[Z \leq 0] = 1/2$ where Z is a standard normal random variable.

Appendix C: FFT and the quadratic estimator

Unfortunately the regression format is not amenable to computation. Therefore one must take advantage of the Fourier filtering characterization of the quadratic estimator to implement it on the computer. There are a few details that we mention here. The nominal form of the quadratic estimator is

$$\hat{\phi}(\mathbf{L}) = A_L \int_{\mathbb{R}^d} \tilde{\Theta}(\ell + \mathbf{L}) \tilde{\Theta}(\ell)^* g_L(\ell) d\ell$$

where the weights g must satisfy the constraint $g_{\mathbf{L}}(\ell)f_{\mathbf{L}}(\ell) \geq 0$ and the normalizing constant $A_{\mathbf{L}}^{-1} \equiv \int g_{\mathbf{L}}(\ell)f_{\mathbf{L}}(\ell)d\ell$ ensure unbiasedness. In [16] the optimal weights g are derived to be proportional to $f_{\mathbf{L}}(\ell)^*/[C_{\mathbf{L}+\mathbf{L},\text{expt}}^{\Theta\Theta}C_{\mathbf{L},\text{expt}}^{\Theta\Theta}]$. However there may be cases where some pairs of frequencies $(\ell, \ell + \mathbf{L})$ need to be excluded. For example, when $\tilde{\Theta}$ is approximated using a discrete FFT one wants to avoid using the frequencies that have a large amount of aliasing: when $|\ell_1| > 2\pi/(2\Delta x_1)$ or $|\ell_2| > 2\pi/(2\Delta x_2)$ (the Nyquist limits, where Δx_1 is the grid spacing in the first coordinate position space). In addition we want to avoid using $\tilde{\Theta}(0)$ (since it contains no information on lensing). To handle this we introduce a masking function $M(\ell)$ and absorb it into the optimal g

$$g_{\mathbf{L}}(\ell) \equiv \frac{f_{\mathbf{L}}(\ell)^*}{C_{\mathbf{L}+\mathbf{L},\text{expt}}^{\Theta\Theta}C_{\mathbf{L},\text{expt}}^{\Theta\Theta}} M(\ell)M(\ell + \mathbf{L}).$$

In our case we use the following mask

$$M(\ell) = \mathbf{1}_{|\ell_1| < \frac{\pi}{2\Delta x_1}} \mathbf{1}_{|\ell_2| < \frac{\pi}{2\Delta x_2}} \mathbf{1}_{\ell \neq 0}$$

to avoid aliasing errors (we cut at half the Nyquist) and the 0 frequency. This masking slightly alters the gradient filtering characterization of the quadratic estimate found in [16]. Under the assumption that the masking function satisfies $M(\ell) = M(-\ell)$ one can derive following characterization of the quadratic estimator:

$$\hat{\phi}(\mathbf{L}) = -2iA_{\mathbf{L}}\mathbf{L} \cdot \int e^{-i\mathbf{x}\cdot\mathbf{L}} \vec{G}(\mathbf{x})W(\mathbf{x}) \frac{d\mathbf{x}}{2\pi} \quad (\text{C1})$$

where $\vec{G}(\ell) \equiv \frac{i\ell C_{\ell}^{\Theta\Theta}\varphi(\ell)^*M(\ell)}{C_{\ell,\text{expt}}^{\Theta\Theta}}\tilde{\Theta}(\ell)$ and $W(\ell) \equiv \frac{\varphi(\ell)^*M(\ell)}{C_{\ell,\text{expt}}^{\Theta\Theta}}\tilde{\Theta}(\ell)$.

Since the masking function is not radially symmetric, the normalization factor $A_{\mathbf{L}}$ will not be either. Since a nominal Riemann approximation to the integral is computationally intensive (a two dimensional integral is required for each \mathbf{L}) we derive a Fourier representation similar to (C1) which can utilize a Fast Fourier Transform. The normalizing factor becomes

$$\begin{aligned} A_{\mathbf{L}}^{-1} &\equiv \int g_{\mathbf{L}}(\ell)f_{\mathbf{L}}(\ell)d\ell \\ &= \int \frac{|f_{\mathbf{L}}(\ell)|^2}{C_{\mathbf{L}+\mathbf{L},\text{expt}}^{\Theta\Theta}C_{\mathbf{L},\text{expt}}^{\Theta\Theta}} M(\ell)M(\ell + \mathbf{L})d\ell \\ &= \sum_{k,j=1}^d \frac{L_k L_j}{\pi} \int e^{-i\mathbf{x}\cdot\mathbf{L}} [A(\mathbf{x})C_{k,j}(\mathbf{x}) - B_k(\mathbf{x})B_j(\mathbf{x})] \frac{d\mathbf{x}}{2\pi} \end{aligned}$$

where $A(\ell) \equiv \frac{|\varphi(\ell)|^2 M(\ell)}{C_{\ell,\text{expt}}^{\Theta\Theta}}$, $B_k(\ell) \equiv i\ell_k C_{\ell}^{\Theta\Theta} A(\ell)$ and $C_{k,j}(\ell) \equiv \ell_j \ell_k (C_{\ell}^{\Theta\Theta})^2 A(\ell)$ for $k, j = 1, 2$.

ACKNOWLEDGMENTS

We gratefully acknowledge enlightening discussion with Jim Berger, Lloyd Knox and Alexander van Engelen.

-
- [1] Anderes, E., Knox, L. & van Engelen, A., Phys. Rev D 83, 043523 (2011)
 - [2] Berger, J., Ann. Statist. 4, 223-226 (1976)
 - [3] Berger, J., Ann. Statist. 8, 716-761 (1980)
 - [4] Berger, J., *Statistical Decision Theory and Bayesian Analysis*, Springer (1980)
 - [5] Bucher, M., Carvalho, C. S., Moodley, K., Remazeilles, M., arXiv:1004.3285 (2010)
 - [6] Carvalho, C. S., Moodley, K., Phys. Rev. D 81, 123010 (2010)
 - [7] Dodelson, S., *Modern cosmology*, Academic Press (2003)
 - [8] Eriksen, H. et al., ApJS 155, 227 (2004)
 - [9] Guzik, J., Seljak, U. & Zaldarriaga, M., Phys. Rev. D 62, 043517 (2000)
 - [10] Hanson, D., Challinor, A., Efstathiou, G., Bielewicz, P., arXiv:1008.4403 (2010)
 - [11] Hirata, C. M., Ho, S., Padmanabhan, N., Seljak, U. & Bahcall, N. A., Phys. Rev. D 78, 043520 (2008)
 - [12] Hirata, C., & Seljak, U., Phys. Rev. D 67, 043001 (2003a)
 - [13] Hirata, C., & Seljak, U., Phys. Rev. D 68, 083002 (2003b)
 - [14] Horn, B. 1990, Int'l J. Computer Vision, 5, 37-75
 - [15] Hu, W., Phys. Rev. D 62, 043007 (2000)
 - [16] Hu, W., ApJ 557: L79-L83 (2001)
 - [17] Hu, W., & Okamoto, T., ApJ 574: 566-574 (2002)
 - [18] James, W. & Stein, C., Proc. Fourth Berkeley Symp. Math. Statist. Prob., 1, pp. 361379 (1961)
 - [19] Kamionkowski, M., Kosowsky, A., Stebbins, A., Phys. Rev. D 55, 7368-7388 (1997)
 - [20] Kaplinghat, M., Knox, L., Song, Y., Phys. Rev. Lett. 91, 241301 (2003)
 - [21] Kesden, M., Cooray, A., Kamionkowski, M., Phys. Rev. Lett. 89, 011304 (2002)
 - [22] Kesden, M., Cooray, A., Kamionkowski, M., Phys. Rev. D 67, 123507 (2003)
 - [23] Knox, L., Song, Y., Phys. Rev. Lett. 89, 011303 (2002)
 - [24] Lehmann, E. & J. Romano *Testing Statistical Hypotheses*, Springer (2005)
 - [25] Lehmann, E. & G. Casella *Theory of Point Estimation*, Springer (1998)
 - [26] Lewis, A., Challinor, A. & Lasenby, A., ApJ, 538: 473-476 (2000)
 - [27] Lewis, A. & Challinor, A., Phys. Rep. 429, 1 (2006)
 - [28] Neidinger, R., SIAM Review, Vol. 52, No. 3, pp.545-563 (2010)
 - [29] Okamoto, T., & Hu, W., Phys. Rev. D 67, 083002 (2003)
 - [30] Seljak, U., & Hirata, C., Phys. Rev. D 69, 043005 (2004)
 - [31] Seljak, U., & Zaldarriaga, M., arXiv:astro-ph/9805010 (1998)

- [32] Seljak, U., & Zaldarriaga, M., Phys. Rev. Letters, 82, 13, 2636-2639 (1999)
- [33] Simchony, T., Chellappa, R., & Shao, M. 1990, IEEE Trans. Pattern An. Machine Intell. , 12, 435-446
- [34] Smith, K., Hu W., Manoj, K., Phys. Rev. D 74, 123002 (2006)
- [35] Smith, K., Zahn, O. & Dore, O., Phys. Rev. D 76, 043510 (2007)
- [36] Strawderman, W., Ann. Statist. 42 (1), 385 - 388 (1971)
- [37] Wandelt, B, Larson, D., Lakshminarayanan, A., Phys. Rev. D 70, 083511 (2004)
- [38] Zaldarriaga, M., & Seljak, U., Phys. Rev. D 59, 123507 (1999)
- [39] Zaldarriaga, M., Phys. Rev. D, 62 (2000)